Journal of Forestry Research (2009) 20(4): 293–300
DOI 10.1007/s11676-009-0051-x

RESEARCH PAPER

A system for calculating the merchantable volume of oak trees in the northwest of the state of Chihuahua, Mexico

Marín POMPA-GARCÍA[1], José Javier CORRAL-RIVAS[1*], José Ciro HERNÁNDEZ-DÍAZ[2], Juan Gabriel ALVAREZ-GONZÁLEZ[3]

[1] *Universidad Juárez del Estado de Durango, Río Papaloapan y Blvd. Durango S/N, Col. Valle del Sur, 34120 Durango, México.*

[2] *Instituto de Silvicultura e Industria de la Madera. Universidad Juárez del Estado de Durango. Km 5.5, Carretera a Mazatlán; CP 34120, Durango, Dgo., México.*

[3] *Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela. Escuela Politécnica Superior, Campus universitario, 27002 Lugo, Spain.*

Abstract: The taper functions of Kozak (1988), Bi (2000) and Fang et al. (2000) were comparatively analyzed in the present paper to develop a system for calculating the merchantable volume of oaks in the northwestern region of the state of Chihuahua (Mexico). Taper data corresponding to 298 trees were collected in mixed and uneven-aged pine-oak stands located throughout the study area, and covering the existing range of ages, stand densities and sites. Results show that the compatible segmented model developed by Fang et al. (2000) best described the experimental data and is therefore recommended for estimating tree diameter at a specific height, height to a specific diameter, merchantable volume, and total volume for oaks. The equation developed in this study is a fundamental tool for use in forest surveys in the study region and is simple enough to ensure its operational implementation. The results of the statistical analysis show that the equation can be recommended for other regions, although some local adaptations may be needed.

Keywords: commercial volume; *Quercus*; taper function; variable inflexion point

Introduction

Many efforts have been made to model tree taper and commercial volume. During the last 30 years, studies have ranged from straight forward taper models (Kozak et al. 1969; Ormerod 1973; Hilt 1980; Lee et al. 2003; Hussain et al. 2008) to polynomial equations (Bruce et al. 1968; Max and Burkhart 1976; Cao et al. 1980), hyperbolic models (Wabo et al. 2002), geometric and trigonometric models (Parresol and Tomas 1996; Fang and Bailey 1999; Zhang et al. 2002; Bi 2000), and Spline functions (Castedo 2003). However, Newnham (1988) states that there are at least two reasons to continue studying this topic: 1) There is still no theory that adequately explains the variability in the shape of the stem for all trees, and 2) equations for estimating tree taper constitute an efficient tool for estimating both commercial and total volume.

Oaks (*Quercus*, Fagaceae) are one of the largest and the most important taxonomic groups in Mexico, where almost 200 species of the 400 reported throughout the world occur (Rzedowski 1978). The forest region in the state of Chihuahua comprises an area of approximately 4,598,454 hectares and is very important in terms of the country's timber production. In this region, oaks are the second most important group of species in terms of its commercial value, the distribution range as well as the harvested timber volume (SEMARNAT 2005). Harvesting of oak species provides an important income in the region, where the main economic activity is the social forestry practice.

Diverse species of oaks occur in La Sierra Madre Occidental (western Mexico) and their taxonomy is complex. Moreover, many of the taxa defined by Trelease (1924) as species are actually only forms of several species. Problems related to classification are intensified by the presence of interspecific hybridization and genetic introgression between many species. In the State of Durango alone at least 43 species are known (González and González 1995). Foresters therefore manage oaks as a species group, and estimate harvestable timber volume at the genus level.

Despite the importance of oaks in Mexico, these species have received scant attention in terms of timber production, since they have been considered to be of less commercial interest than pine species. Consequently, there are few studies regarding methods of determining stem standing volume (e.g. Pompa et al. 1998). Most studies carried out to test the suitability of different taper

Foundation project: This study was supported by by El Fondo de Cooperación Internacional en Ciencia y Tecnología entre la UE-México (project: 92739)

Received: 2009-05-18; Accepted: 2009-07-22

The online version is available at http://www.springerlink.com

Biography: José Javier Corral-rivas (1974-), male, Dr., Professor in the Department of Forest Management, Faculty of Forest Sciences, Juarez University of Durango, Mexico. E-mail: jcorral@ujed.mx

Responsible editor: Hu Yanbo

functions for describing the stem profile of different tree species have been performed with pines (Návar and Domínguez 1997; Návar et al. 1997; Corral-Rivas et al. 2007). These studies are very useful for both forest research and forestry practice, and may help in attaining sustainable forest management of this timber resource. To date, no studies have been conducted to develop merchantable volume equations for oaks in the study area. The objective of this research was therefore to develop a merchantable volume system for oaks (*Quercus*, Fagaceae) in the northwestern region of the State of Chihuahua in Mexico. The analyses involved the use of tenable statistical assumptions for accounting for the main problems associated with the construction of taper functions and volume equations (multicollinearity, autocorrelated errors and heteroscedasticity), while examining the residuals associated with each position and diameter classes.

Materials and methods

Study site

The study was conducted in the northwestern region of the state of Chihuahua (Mexico), in an area of 251 960 ha between longitudes of 108˚15' and 108˚45' West and latitudes of 28˚45' and 30˚00' North (Fig. 1). The study area includes parts of the municipalities of Madera and Casas Grandes, in the high range of the Sierra Madre Occidental, where the topography is variable, with irregular elevations and depressions. The predominant forest types are pine-oak stands, often mixed with *Pseudotsuga* and *Juniperus,* among other species (SEMARNAT 2005).

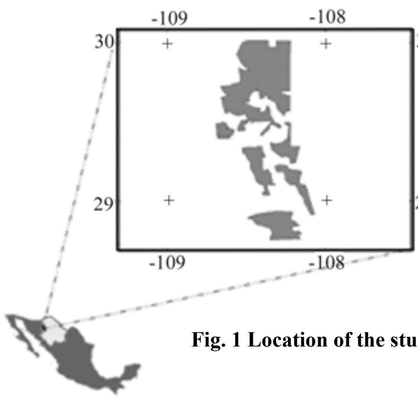

Fig. 1 Location of the study area

Description of data

Taper data corresponding to 298 trees were collected in mixed and uneven-aged stands located throughout the study area, and covering the existing range of ages, stand densities and sites. The trees were subjectively selected to ensure a representative distribution of diameter classes. The diameter at breast height (*D*, in cm) and the total height (*H*, in m) were measured in each tree. The trees were cut into logs at 1 to 3 m intervals along the stem. Two perpendicular diameters outside-bark were measured in each section, and then averaged. Log volumes were calculated in cubic meters, with Smalian's formula. The top section was treated as a cone. Outside-bark total stem volume (above stump) was obtained by summing the outside-bark log volumes and the volume of the top of the tree.

Summary statistics of the final data set used in this study are shown in Table 1. Plots of relative height (*h/H*) against relative diameter (*d/D*) for each species are shown in Fig. 2. The width of the distribution of the data reflects the size and shape of the trees constituting the sample.

The following notation will be used hereafter:

D= diameter at breast height over bark (1.3 m above ground, cm);

d= diameter over bark at height h (cm);

H= total tree height (m);

h= height above ground to diameter d (m);

hst= stump height (m);

V= total stem volume under bark from stump (m^3);

v= merchantable volume under bark (m^3), the volume from stump to the point where diameter = d;

a_i, b_i, p_i= coefficients to be estimated through the adjustment;

k=π/40000, a metric constant for converting from squared diameter, in cm^2, to cross-section area, in m^2;

q= h/H

Table 1. Summary statistics for the sampled trees

Variable	No. of Observations	Media	Standard deviation	Maximum	Minimum
D (cm)	298	40.9	17.1	81.5	8.2
D (cm)	1646	25.2	19.5	109.5	0
H (m)	298	14.9	5.7	34.75	4.0
H (m)	1646	7.1	5.6	34.75	0.1

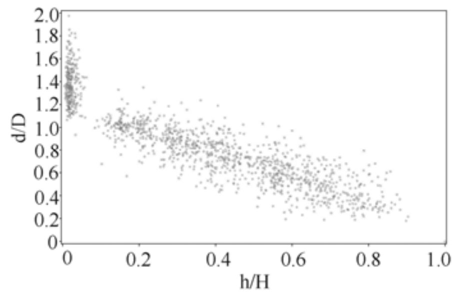

Fig. 2 Plot of relative diameter against relative height for the 298 trees used to fit the taper equations

Functions selected for comparison

In the present study, the taper functions of Kozak (1988), Bi (2000) and Fang et al. (2000) were fitted to the data set. These functions provided good results in other applications and have shown great flexibility in fitting variable stem forms (Diéguez-Aranda et al. 2006; Corral-Rivas et al. 2007). The functions can be classified into two groups: variable-form taper models and segmented taper models.

A variable-form taper model describes the stem shape with a changing exponent or variable from ground to top to represent

the neiloid, paraboloid, conic and several intermediate forms (Kozak 1988; Newnham 1988). This approach is based on the assumption that the stem form varies continuously along the length of a tree (Lee et al. 2003). In comparison with single and segmented taper models, this approach usually provides the lowest degree of local bias and greatest precision in taper predictions (e.g., Kozak 1988; Pérez et al. 1990; Newnham 1992; Muhairwe 1999; Rojo et al. 2005). However, the disadvantage is that they cannot be integrated analytically to calculate total stem or log volumes. In the present study, we analyzed the models proposed by Kozak (1988) and Bi (2000):

Kozak (1988) taper function:

$$d = b_1 D^{b_2} b_3{}^{D} \left(\frac{1-\sqrt{q}}{1-\sqrt{q}} \right)^{\left(\begin{array}{c} b_4 q^2 + b_5 Ln(q+0{,}001) + \\ b_6\sqrt{q} + b_7 e^{q} + b_8\left(\frac{D}{h}\right) \end{array} \right)} \quad (1)$$

Bi (2000) taper function:

$$\frac{d}{D} = \left(\frac{Ln\sin\left(\frac{\pi q}{2}\right)}{Ln\sin\left(\frac{\pi 1{,}3}{2h}\right)} \right)^{\left(\begin{array}{c} b_1 + b_2 sen\left(\frac{\pi q}{2}\right) + b_3\cos\left(\frac{3\pi q}{2}\right) + \\ b_4 \frac{sen\left(\frac{\pi q}{2}\right)^2}{q} + b_5 D + b_6 q\sqrt{D} + b_7 q\sqrt{h} \end{array} \right)} \quad (2)$$

In describing tree taper, it is generally accepted that a tree stem can be divided into three geometric shapes: the top is considered as a cone, the central section a frustum of a paraboloid, and the butt a frustum of a neiloid. Segmented models describe these shapes by fitting each with an equation, and then mathematically joining the segments to produce an overall segmented function. From this group of segmented taper models, we analyzed the compatible model developed by Fang et al. (2000).

Fang et al. (2000) taper function:

$$d = c_1\sqrt{H^{(k-b_1)/b_1}(1-q)^{(k-\beta)/\beta}\alpha_1^{I_1+I_2}\alpha_2^{I_2}} \quad (3)$$

Where

$$c_1 = \sqrt{\frac{a_0 D^{a_1} H^{a_2 - k/b_1}}{b_1(r_0 - r_1) + b_2(r_1 - \alpha_1 r_2) + b_3\alpha_1 r_2}};$$

$$\beta = b_1^{1-(I_1+I_2)} b_2^{I_1} b_3^{I_2}; \quad \alpha_1 = (1-p_1)^{\frac{(b_2-b_1)k}{b_1 b_2}};$$

$$\alpha_2 = (1-p_2)^{\frac{(b_3-b_2)k}{b_2 b_3}};$$

$$r_0 = (1-h_{st}/H)^{k/b_1}; \quad r_1 = (1-p_1)^{k/b_1};$$

$$r_2 = (1-p_2)^{k/b_2}$$

$$\begin{cases} I_1 = 1 \text{ if } p_1 \le q \le p_2; \text{ 0 otherwise} \\ I_2 = 1 \text{ if } p_2 < q \le 1; \text{ 0 otherwise} \end{cases}$$

$p_1 = h_1/H$ and $p_2 = h_2/H$ (relative heights from the ground up to the sections where the two union points assumed in the model occur; the first is close to the normal height (1.3 m) and the second occurs in a higher section of the stem).

Fang et al. (2000) also derived a compatible model for merchantable v and total V volume by direct integration of the taper model. Their expressions are as follows:

$$v = c_1^2 H^{k/b_1}\left(\begin{array}{l} b_1 r_0 + (I_1+I_2)(b_2-b_1)r_1 + \\ I_2(b_3-b_2)\alpha_1 r_2 - \beta(1-q)^{k/\beta}\alpha_1^{I_1+I_2}\alpha_2^{I_2} \end{array} \right) \quad (4)$$

$$V = a_0 D^{a_1} H^{a_2} \quad (5)$$

Although the development of the compatible system of Fang et al. (2000) is based on Eq. (5), any other volume equation can be used as input into the system.

Model fitting

To avoid problems in the estimation of the parameters, especially when $h = H$, i.e. when $d = 0$, a small value lower than the measurement limit used in the data collection was reassigned to those diameters equal to zero. Similarly, a value lower than the corresponding height measurement limit was subtracted from the heights equal to the total height. A similar approach was used by Fang et al. (2000) in order to avoid the logarithm of zero in model fitting. This approach allows the use of the entire dataset for fitting and does not significantly change the parameter estimates (Fang et al. 2000; Diéguez-Aranda et al. 2006; Corral-Rivas et al. 2007).

Estimation of the parameters was carried out with the MODEL procedure of SAS/ETS® (SAS Institute Inc. 2004), in which several methods for parameter estimation are available. For the models of Kozak (1988) and Bi (2000) we used generalized nonlinear least squares, since the models are composed of only one equation. The compatible system of Fang et al. (2000) has two components: a taper function and a volume equation to any specified diameter limit, in which the total volume equation is a special case included in the merchantable volume equation (i.e., when $h = H$, then $V = f(D,H)$). We used the seemingly unrelated regression technique to fit the system of Fang et al. (2000), in which the random errors of the equations are correlated, but the equations are not really simultaneous (none of the endogenous variables in one equation of the system appears as dependent on the left-hand side of the other equation; Zellner 1962; Judge et al. 1988; Rose and Lynch 2001). This approach requires estimation of the parameters of the taper function and recovery of the implied total volume equation.

Multicollinearity, autocorrelation and heteroscedasticity

There are several problems associated with stem taper and volume equation analysis that violate the fundamental least squares assumptions of independence and equal distribution of errors with zero mean and constant variance, of which multicollinearity, autocorrelation and heteroscedasticity are three of the most important. Although the least squares estimates of regression coef-

ficients remain unbiased and consistent in the presence of multicollinearity, autocorrelation and heteroscedasticity, they are no longer efficient (Myers 1990). These problems may seriously affect the standard errors of the coefficients, thereby invalidating statistical tests that use t or F distributions and confidence intervals (Neter et al. 1990, p.300). Thus, appropriate statistical procedures should be used in model fitting to avoid problems associated with heteroscedasticity and autocorrelated errors, and models with low multicollinearity should be selected whenever possible (Kozak 1997).

To evaluate the presence of multicollinearity among variables in the models analyzed, we used the condition number, which is defined as the square root of the ratio of the largest to the smallest eigenvalue of the correlation matrix. According to Belsey (1991), if the condition number is between 5 and 10, collinearity is not a major problem, if it is in the range of 30–100, then there are problems associated with collinearity, and if it is in the range 1 000–3 000 there are severe problems associated with collinearity. Since the database contains multiple observations for each tree (i.e., hierarchical data), it is reasonable to expect that the observations within each tree are spatially correlated, which violates the assumption of independent error terms. We used a continuous-time autoregressive error structure (CAR(x)) to account for the inherent autocorrelation of the data. This error structure enables the model to be applied to irregularly spaced, unbalanced data (Zimmerman and Nuñez-Antón 2001; Diéguez-Aranda et al. 2006; Corral-Rivas et al. 2007). To test for the presence of autocorrelation and the order of the CAR(x) to be used, graphs of residuals plotted against lag-residuals from previous observations within each tree were examined visually. Appropriate fits for the models with correlated errors were made by including the CAR(x) error structure in the MODEL procedure of SAS/ETS®system (SAS Institute Inc. 2004), which allows for dynamic updating of the residuals.

Model comparison

The criteria used for judging the performance of the taper functions were based on numerical and graphical analyses of the residuals. Three goodness-of-fit statistics were examined: bias ($\overline{E}$), coefficient of determination for nonlinear regression (R^2) (see Ryan 1997, pp. 419 and 424), root mean square error (*RMSE*) and the Akaike information criterion in differences *(ΔAIC)*, which is an index used to select the best model and is based on minimizing the distance of Kullback-Liebler. Although there are several shortcomings associated with the use of the R^2 in nonlinear regression, the general usefulness of some global measure of model adequacy would appear to override some of these limitations (Ryan 1997, p. 424). The expressions of these fitting statistics are summarized as follows:

$$\overline{E} = \sum_{i=1}^{n}(y_i - \hat{y}_i)/n \tag{7}$$

$$R^2 = r^2_{Y_i\hat{Y}_i} \tag{8}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p)} \tag{9}$$

$$\Delta\mathrm{AIC} = n\cdot\ln\hat{\sigma}^2 + 2\cdot(p+1) - \min(n\cdot\ln\hat{\sigma}^2 + 2(p+1))$$
$$\text{with } \hat{\sigma}^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2/n \tag{10}$$

where $r_{Y_i\hat{Y}_i}$ is the coefficient of correlation between the measured (y_i) and estimated ($\hat{y}_i$) values of the dependent variable, n is the total number of observations used to fit the model, p is the number of model parameters, and $\hat{\sigma}^2$ the estimator of the variance of the model error.

Because the quality of fit does not necessarily reflect the quality of future prediction (Myers 1990, p. 168), assessment of the validity of the model with an independent data set is recommended (Kozak and Kozak 2003). Due to the scarcity of such data, several methods have been proposed (e.g., splitting the data set or cross-validation, double cross-validation), although they seldom provide any additional information compared with the respective statistics obtained directly from models built from entire data sets (Kozak and Kozak 2003). Thus, because decisions have to be made with available information, it is better to wait to obtain new data before such validation is carried out (Diéguez-Aranda et al. 2006; Corral-Rivas et al. 2007).

The taper functions were also assessed by use of box plots for d residuals by position (percent relative height points along the stem, i.e., 5%, 15%, 25%, and so on up to 95%). We did the same for h residuals by d classes and for V residuals by D classes. These graphs, calculated by position or diameter class, are very important for showing areas or diameter classes for which the functions provide especially poor or good predictions.

Results and discussion

Initially, the models were fitted by use of nonlinear least squares without expanding the error terms to account for autocorrelation. A trend in residuals as a function of lag1- and lag2-residuals within the same tree was apparent in all the models analyzed, as expected because of the longitudinal nature of the data used for model fitting. Figure 3 (first row) provides an example of this with the model of Fang et al. (2000). After correcting for autocorrelation with a second-order continuous-time autoregressive error structure, the trends in the residuals disappeared (Fig. 3 third row). No higher-order autocorrelation was observed (Fig. 3, third column). The estimations provided by the model were not significantly different from the models fitted without considering such correction. The sole purpose of autocorrelation correction was to improve interpretation of the statistical properties of the model, and it has no practical application unless several measurements of diameter at different heights are considered in the same individual tree.

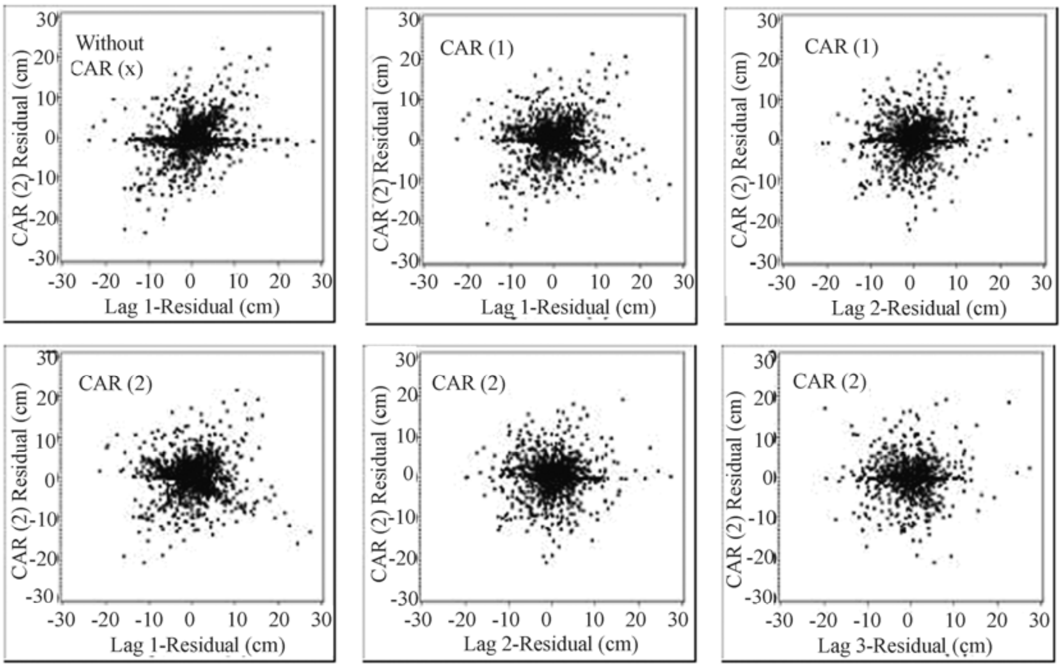

Fig. 3 Graphs of *d* residuals plotted against: Lag1-Residuals (left column), Lag2-Residuals (middle column), and Lag3-Residuals (right column) for the model of Fang et al. (2000) fitted without considering the autocorrelation parameters (first row), and with continuous-time first and second order autoregressive error structures (second and third rows, respectively)

Table 2. Parameter estimates, approximated standard errors, goodness-of-fit statistics and condition number of the models analyzed

Model	Parameter	Estimate	Standard error	Goodness-of-fit statistics R^2	RMSE	E	ΔAIC	Condition number
Kozak (1988)	b_1	1.5584	0.2024					
	b_2	0.7836	0.0466					
	b_3	1.0020	0.0011					
	b_4	0.3885	0.1170	0.94	4.73	-0.17	106.12	130.71
	b_5	-0.3027	0.0276					
	b_6	2.5502	0.3143					
	b_7	-0.9362	0.1575					
	b_8	0.0145	0.0044					
Bi (2000)	b_1	4.0586	0.1365					
	b_2	-0.7308	0.0456					
	b_3	n.s						
	b_4	-2.3205	0.0844	0.93	5.44	0.24	636.71	97.66
	b_5	n.s						
	b_6	0.0298	0.0096					
	b_7	-0.058	0.0139					
Fang et al. (2000)	a_0	0.000076	6.71×10^{-6}					
	a_1	1.8172	0.026					
	a_2	1.0014	0.0246					
	b_1	6.27×10^{-6}	3.96×10^{-7}					
	b_2	3.1×10^{-6}	4.68×10^{-7}	0.95	4.59	-0.04	0	35.75
	b_3	3.9×10^{-5}	1.7×10-6					
	p_1	0.0556	0.0031					
	p_2	0.5992	0.0331					

n.s = not significant

Table 2 shows the parameter estimates and their corresponding approximated standard errors along with the goodness-of-fit statistics. All the parameters were found to be significant at the 1% significance level, except for parameters b_3 and b_5 in the model of Bi (2000), which were excluded from the model because the null hypothesis could not be rejected, which suggests that the values are equal to zero. The multicollinearity of the models was moderate, as inferred from the condition numbers, which had values of 37.75, 97.66, and 130.71 for the models of Fang et al. (2000), Bi (2000) and Kozak (1988), respectively. All models accounted for more than 93% of the total variance of *d* along the stem, particularly the function of Fang et al. (2000), which explained more than 95% of this variability (see the R^2 values in Table 2). The values of the mean square error were 4.7, 5.4 cm, and 4.6, for the models of Kozak (1988), Bi (2000), and Fang et al. (2000), respectively. According to these goodness-of-fit statistics, the model of Fang et al. (2000) best described the data for diameter to a specific height. Moreover this function showed the smallest multicollinearity as inferred from the condition number.

The box plots of *d*, and *h* residuals against relative heights and *D* classes, respectively (Fig. 4) showed a non-homogeneous distribution of the residuals for the variables *d* and *h* for the model of Fang et al. (2000). However, some unsatisfactory estimations were observed in all zones, although the mean was not skewed and the 25% and 75% quartiles of the distribution of residuals were very close to zero. At the bottom of the stem, where most of the volume is accumulated and where greater irregularities are usually present, especially in oaks, the residuals in diameter estimation are higher. This is caused by the lack of observations below a height of 0.5 m from the stump, which would improve the fit in this zone.

In summary, the model of Fang et al. (2000) best described the data for both diameter along the stem and height to a specific diameter. Thus, the compatible volume system of Fang et al. (2000) is proposed as the most suitable for describing the stem profile and predicting stem volume of oaks in the northwest of

the State of Chihuahua.

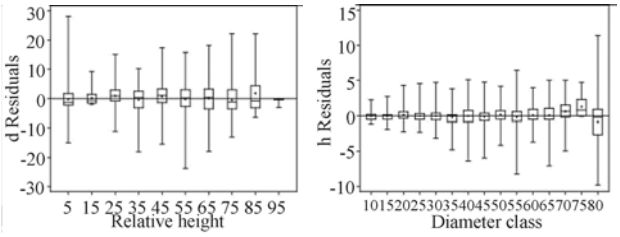

Fig. 4 Box plots of *d* residuals (Y-axis, in cm) against relative height classes (X-axis, in percent) (left) and of *h* residuals (Y-axis, in m) against *D* classes (X-axis, in cm) (right) for the model of Fang et al. (2000). The plus signs represent the mean of prediction errors for the corresponding relative height and diameter at breast height classes, respectively. The boxes represent the interquartile ranges. The maximum and minimum diameter under bark and height prediction errors are represented by the small horizontal lines at the upper and lower ends of the vertical lines, respectively

Comparison of the taper function for different diameter classes

One of the main objectives of this type of study is to obtain taper functions that are valid for a wide range of site indexes, densities and characteristics of the trees in a stand. Good discussions of the factors affecting the shape and profile of the trees are provided by Larson (1963) and LeMay et al. (1991). There are many studies in which attempts have been made to relate the parameters of taper functions to characteristics of the site index or to the size of the tree. For example, Morris and Forslund (1992) observed that the microsite and climatic variables explained 61.9% and 38.7% of the variation in the taper function and the tree shape, respectively, with the model of Forslund applied to *Pinus banksiana*. Muhairwe et al. (1994) found that inclusion of the age, site index and/or variables related to the tree crown in the model of Kozak (1988), slightly improved the estimations obtained for *Pseudotsuga menziesii*, *Populus tremuloides* and *Thuja plicata*.

Research efforts directed at developing taper functions that accounted for the effect that these factors have on the differences observed in the tape along the tree stem gave rise to, amongst others, the complete variable exponent models (e.g. Kozak 1988 or Bi 2000) or segmented models (Fang et al. 2000).

The size of a tree is another factor that has great effect in determining the taper of the tree. Bi and Hamilton (1998) observed the importance of the dimensions of the tree to its profile. More recently, in a study with several species of Eucalyptus, Bi (2000) showed that tree size affected the value of the inflexion point of its estimated profile function, which made him consider that the models that force the use of a common inflexion point for all trees are less flexible. Similar results were obtained by Bi and Long (2001) for *Pinus radiata*.

In the model of Fang et al. (2000), two inflexion points (p_1 and p_2), which are common to all of the trees, were used. The authors consider that two inflexion points are sufficient for most species, in detail, the first would be located close to the diameter at breast height (1.30 m) and the second one would be further up the stem. We therefore proceeded to analyze the effect of tree size, represented by the diameter class, at such inflexion points.

The first step was to group the trees in diameter classes of 10 cm. The number of trees in each diameter class, as well as the mean, maximum and minimum values and the typical deviations in the heights and normal diameters are shown in Table 3. The number of trees in the higher diameter classes is very small and therefore all the trees with diameter larger than 60 cm were grouped in a single class.

Table 3. Mean, minimum and maximum values, typical deviations in heights and diameters, and number of trees in each diameter class

Class interval (cm)	Trees (No.)	Diameter (cm)				Heights (m)			
		Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
< 20	57	14.86	3.0082	8.20	19.60	8.71	3.1455	4.07	18.72
20-30	59	25.03	3.0684	20.00	29.70	10.53	3.7542	4.61	18.83
30-40	56	34.93	2.9821	30.00	39.90	13.60	4.1000	5.34	22.80
40-50	46	45.01	2.8617	40.00	49.80	15.36	3.6282	7.74	21.95
50-60	47	54.81	2.8343	50.00	59.40	15.58	5.5831	5.31	25.69
60-70	25	63.22	2.3056	60.00	68.70	16.44	5.8764	8.10	28.26
70-80	5	74.46	3.2145	70.00	77.50	22.15	6.5013	13.72	30.60
80-90	3	81.23	0.2517	81.00	81.50	29.21	7.5715	20.58	34.75

To assess whether the taper equation differs among diameter classes, the nonlinear extra sum of squares method was used (Bates and Watts 1988, pp. 103-104). This method requires the fitting of full and reduced models, and has frequently been applied to assess whether separate models are necessary for different species or different geographic regions (e.g., Huang et al. 2000; Zhang et al. 2002; Corral-Rivas et al. 2004; Corral-Rivas et al. 2005; Castedo et al. 2005). In this case, the reduced model corresponds to the same set of global parameters for all diameter classes, while the full model corresponds to different sets of global parameters for the inflexion points (p_1 and p_2) of the model in each diameter class. The full model is obtained by expanding each global parameter by including an associated parameter and a dummy variable to differentiate the diameter classes.

$$p_i + c_{ik} I_k \qquad i = 1,2 \quad k = 2,...,6 \tag{11}$$

where p_i is one of the inflexion points of the model of Fang et al. (2000), c_{ik} is a parameter associated with the inflexion point i and the diameter class k in the full model, and I_k is a dummy variable whose value is equal to 1 for the diameter class k and 0 for the other classes. The appropriate test statistic uses the following expression:

$$F^* = \left(\frac{SSE_R - SSE_F}{df_R - df_F} \right) \frac{df_F}{SSE_F} \tag{12}$$

where SSE_R is the error sum of squares of the reduced model, SSE_F is the error sum of squares of the full model, and df_R and df_F are the degrees of freedom of the full and reduced models, respectively. The non-linear extra sum of squares follows an F-distribution.

Table 4. Parameter estimates, approximated standard errors, goodness-of-fit statistics and condition number of the model of Fang et al. (2000), fitted with the same inflexion points for each diameter class (reduced model) and with a different inflexion point per diameter class (full model)

Model	Parameter	Estimate	Standard error	Goodness-of-fit statistics R^2	RMSE	$\bar{E}$	ΔAIC	Condition Number
Fang et al.(2000) Reduced model	a_0	0.000076	6.71×10^{-6}	0.95	4.59	-0.04	28.53	35.75
	a_1	1.8172	0.026					
	a_2	1.0014	0.0246					
	b_1	6.27×10^{-6}	3.96×10^{-7}					
	b_2	3.1×10^{-6}	4.68×10^{-7}					
	b_3	3.9×10^{-5}	1.7×10^{-6}					
	p_1	0.0556	0.0031					
	p_2	0.5992	0.0331					
Fang et al. (2000) Full model	a_0	0.000063	6.57×10^{-6}	0.95	4.56	-0.00	0	45.01
	a_1	1.8675	0.0301					
	a_2	0.9944	0.0252					
	b_1	6.08×10^{-6}	4.04×10^{-7}					
	b_2	3.1×10^{-5}	4.71×10^{-7}					
	b_3	3.8×10^{-5}	1.69×10^{-6}					
	p_{11}	0.0667	0.0046					
	p_{13}	-0.0118	0.0035					
	p_{14}	-0.0122	0.0034					
	p_{15}	-0.0182	0.0035					
	p_{16}	-0.0108	0.0035					
	p_{21}	0.5975	0.0341					

Results of the fitting process for full and reduced forms of the model of Fang et al. (2000) are shown in Table 4. The F statistic calculated by Eq. (12) was 4.3, and the probability of finding a critical value greater than 4.3 (i.e., F critical (1-α; df_R - df_F) > 4.3) was lower than 0.01. Differences among the taper equations were therefore observed for different diameter classes. This means that the full model better represents the general tendency of the sample trees. The diameter classes have an important effect on the value of the first inflexion point, i.e., the inflexion point located around the diameter at breast height (all the parameters associated with each diameter class are significant except for parameter p_{12}; see Table 4). However, the size of the trees, represented by the diameter classes, has no effect on the position of the second inflexion point along the stem, since none of the estimators of the parameters associated with each diameter class were significant ($p_{22},\ldots,p_{26}$), i.e., all the diameter classes can be well represented with a single value (p_{21}).

The autocorrelation for the full model was also corrected by use of a second-order continuous-time autoregressive error structure and as in the reduced model, the trends in residuals disappeared. The box plots of d, and h residuals against relative heights and D classes, respectively were very similar to those calculated for the reduced model (see Fig. 4) and showed a non-homogeneous distribution of the residuals for the variables d and h.

Conclusions

Among the three taper functions assessed for describing the taper for oaks species in the northwestern region of the State of Chihuahua (Mexico), the segmented compatible model of Fang et al. (2000) best described the experimental data. This model performed well in predicting upper-stem diameter, height to any specific diameter and total volume. The system of Fang et al. (2000) also has the advantage that the resulting taper, merchantable and total volume equations are compatible with each other. The non-linear extra sum of squares method indicated significant differences in the estimator of the inferior inflexion point in the stem profile (i.e. that located around the breast height), but not for the second inflexion point. In this sense, Hernández (2004) obtained results indicating that the inflexion points among the stem shapes decrease over time, in other words, taper in early growth classes differs from the taper presented by older individuals. Therefore, separation of diameter classes is recommended for estimating upper-stem diameter, height to any specific diameter and total volume, especially because Mexican oaks usually have unstable stem shapes. In this respect, Castedo (2003) pointed out that such modelling requires further research, since the exact number of inflexion points is still unknown. Spline functions may be useful for such purposes. The equation developed in the present study is a fundamental tool for application in forest surveys in the study region; it is simple enough to ensure its operational implementation. Moreover the statistical shows that its implementation in other similar regions can be recommended, although some local adaptations may be needed.

Acknowledgments
We acknowledge the support offered by the technical direction of the Ejido “El Largo y Anexos” in providing the necessary data for this study. This research was also generously supported by El Fondo de Cooperación Internacional en Ciencia y Tecnología entre la UE-México (project: 92739).

References

Bates DM, Watts DG. 1988. *Nonlinear Regression Analysis and Its Applications*. Wiley, New York, 365 pp.

Belsey DA. 1991. *Conditioning diagnostics, collinearity and weak data in*

regression. John Wiley & Sons, Inc., New York, 396 pp.
Bi H. 2000. Trigonometric variable-form taper equations for Australian eucalyptus. *For Sci,* **46**: 397–409.
Bi H, Hamilton F. 1998. Stem volume equations for native tree species in southern New South Wales and Victoria. *Aus For*, **61**: 275–286.
Bi H, Long Y. 2001. Flexible taper equation for site-specific management of *Pinus radiata* in New South Wales, Australia. *For Ecol Manage*, **148**: 79–91.
Bruce DR, Curtis O, Vancoevering C. 1968. Development of system and taper volume tables for red alder. *For Sci*, **14**: 339–350.
Cao QV, Burkhart HE, Max TA. 1980. Evaluations of two methods for cubic-foot volume prediction of loblolly pine to any merchantable limit. *For Sci*, **26**: 71–80.
Castedo DF. 2003. Modelo dinámico de crecimiento para las masas de *Pinus radiata* D. Don en Galicia. Tesis doctoral. Universidad de Santiago de Compostela, España, 297 pp.
Castedo F, Barrio M, Parresol BR, Álvarez JG. 2005. A stochastic height-diameter model for maritime pine ecoregions in Galicia (northwestern Spain). *Ann For Sci*, **62**: 455–465.
Corral-Rivas JJ, Álvarez JG, Ruíz GA, Gadow Kv. 2004. Compatible height and site index models for five pine species in El Salto, Durango (Mexico). *For Ecol Manage*, **201**: 145–160.
Corral-Rivas JJ, Álvarez GJ, Aguirre CO, Hernández FJ. 2005. The effect of competition on individual tree basal area growth in mature stands of *Pinus cooperi* Blanco in Durango (Mexico). *Eur J Forest Res,* **124**: 133–142.
Corral-Rivas JJ, Diéguez-Aranda U, Corral S, Castedo F. 2007. A merchantable volume system for major pine species in El Salto, Durango (Mexico). *For Ecol Manage*, **238** (1-3): 118–129.
Diéguez-Aranda U, Castedo-Dorado F, Álvarez JG, Rojo A. 2006. Compatible taper function for Scots pine plantations in northwestern Spain. *Can J For Res,* **36**(5): 1190–1205.
Fang Z, Bailey RL. 1999. Compatible volume and taper models with coefficients for tropical species on Hainan Island in Southern China. *For Sci,* **45**: 85–100.
Fang Z, Borders BE, Bailey RL. 2000. Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. *For Sci*, **46**: 1–12.
González EMS, González ES. 1995. Los encinos de Durango, México. En: J. S. Marroquín de la F. (De.). III Seminario nacional sobre utilización de encinos. Reporte Científico Número Especial 15. Fac. Cienc. Ftales. UANL. Linares, N.L. México. Tomo I: 234–237.
Hilt DE. 1980. Development of taper functions from variable-top merchantable volume equations. *For Sci*, **26**: 117–120.
Hernández ER. 2004. Función de ahusamiento y simulador de trozado para *Sequoia sempervirens* (D.Don) Endl. 2004. Tesis de ingeniero forestal. Facultad de ciencias forestales. Universidad de Chile, 84 pp.
Huang S, Price D, Titus SJ. 2000. Development of ecoregion-based height-diameter models for white spruce in boreal forests. *For Ecol Manage*, **129**: 125–141.
Hussein KA, Schmidt M, Kotzé H, Gadow Kv. 2008. Parameter-parsimonious Taper Functions for Describing Stem Profiles. *Scientia Silvae Sinicae*, **44**(6): 1–8.
Judge GG, Griffiths WE, Hill RC, Lütkephol H, Lee TC. 1988. *Introduction to the theory and practice of econometrics*. John Wiley & Sons, New York, 1024 pp.
Kozak A. 1988. A variable-exponent taper equation. *Can J For Res*, **18**: 1363–1368.
Kozak A. 1997. Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. *Can J For Res*, **27**: 619–629.
Kozak A, Kozak RA. 2003. Does cross validation provide additional information in the evaluation of regression models? *Can J For Res*, **33**: 976–987.
Kozak A, Munro D, Smith J. 1969. Taper functions and their application in forest inventory. *For Chron*, **45**: 278–283.
Larson PR. 1963. Stem form development of forest trees. *For Sci Monogr* 5, pp. 1–41.
Lee WK, Seo JH, Son YM, Lee KH, Gadow Kv. 2003. Modeling stem profiles for *Pinus densiflora* in Korea. *For Ecol Manage*, **172**: 69–77.
LeMay VM, Kozak A, Marshall PL, Muhairwe C. 1991. Literature review for development of a dynamic taper model for tree growth, Internal Report for the B.C. Science Council, Vancouver, British Columbia.
Max TA, Burkhart HE. 1976. Segmented polynomial regression applied to taper equations. *For Sci*, **22**(3): 283–289.
Morris DM, Forslund RR. 1992. The relative importance of competition, microsite, and climate in controlling the stem taper and profile shape in jack pine. *Can J For Res*, **22**: 1999–2003.
Muhairwe CK. 1999. Taper equations for *Eucalyptus pilularis* and *Eucalyptus grandis* for the north coast in New South Wales, Australia. *For Ecol Manage*, **113**: 251–269.
Muhairwe CK, LeMay VM, Kozak A. 1994. Effects of adding tree, stand, and site variables to Kozak's variable-exponent taper equation. *Can J For Res*, **24**: 252–259.
Myers RH. 1990. *Classical and modern regression with applications*. 2nd edition. Duxbury Press, Belmont, California, 488 pp.
Návar J, Domínguez PA. 1997. Ajuste de modelos de volumen y funciones que describen el perfil diamétrico de cuatro especies de pino plantadas en el Nordeste de México. *Invest Agrar Sist Recur For*, **6**(1-2): 147–163.
Návar J, Domínguez PA, Contreras JC Estrada MC. 1997. Ajuste de siete modelos de ahusamiento a los perfiles fustales de *Pinus hartwegii* Lindl. Del Noroeste de México. *Agrociencia*, **31**: 73–81.
Neter J, Wasserman W, Kutner MH. 1990. *Applied linear statistical models: regression, analysis of variance and experimental designs*. 3th edition. Irwin, Boston, 667 pp.
Newnham R. 1988. A variable-form taper function. Petawawa Nt. For. *Inst. Information Rep*, PI-X-83, pp. 1-33.
Newnham R. 1992. Variable-form taper functions for four Alberta tree species. *Can J For Res*, **22**: 210–223.
Ormerod DW. 1973. A simple bole model. *For Chron*, **49**: 136–138.
Parresol BR, Tomas CC. 1996. A simultaneous density-integral system for estimating stem profile and biomass: Slash pine and Willow Oak. *Can J For Res*, **26**: 773–781.
Pérez D, Burkhart HE, Stiff C. 1990. A variable-form taper function for *Pinus oocarpa* Schiede. in Central Honduras. *For Sci*, **36**: 186–191.
Pompa GM, Velázquez MA, Ramírez MH, Rodríguez FC. 1998. Sistema de cubicación para el género *Pinus* en La Ucodefo 7 "Norogachi-Guachochi Chihuahua". *Rev Cienc For*, **23**(83): 17–23.
Rose CE, Lynch, TB. 2001. Estimating parameters for tree basal area growth with a system of equations and seemingly unrelated regressions. *For Ecol Manage*, **148**: 51–61.
Rojo A, Perales X, Sanchez-Rodriguez F, Álvarez JG, Gadow Kv. 2005. Stem taper functions for maritime pine (*Pinus pinaster* Ait.) in Galicia (Northwestern Spain). *Eur J Forest Res*, **124**: 177–186.
Ryan TP. 1997. Modern regression methods. John Wiley & Sons, New York, 515 pp.
Rzedowski J. 1978. Vegetación de México. Editorial Limusa. México. 432 pp.
SAS Institute Inc. 2004. SAS/ETS® 9.1 User's Guide. Cary, NC: SAS Institute Inc.
SEMARNAT. 2005. Control técnico de aprovechamientos. Listado electrónico. Delegación federal de Chihuahua. Subdelegación de gestión para la protección ambiental y recursos naturales, 150 pp.
Trelease W. 1924. The American Oaks. *Prc National Acad Sci* (US), **20**: 1–255.
Wabo E, Marquina J, Bratovich R. 2002. Modelos de perfil de fuste para clones comerciales de *Populus deltoides* Marsh en la zona continental de la provincia de Buenos Aires. *Quebracho Rev Cienc For*, **9**: 83–94.
Zellner A. 1962. An efficient method of estimating seemingly unrelated regressions and test for aggregation bias. *J Am Stat Assoc*, **57**: 348–368.
Zhang L, Peng L, Huang S, Zhou X. 2002. Development and evaluation of ecoregion-based tree height-diameter models for jack pine in Ontario. *For Chron*, **78**: 530–538.
Zimmerman DL, Núñez-Antón V. 2001. Parametric modelling of growth curve data: An overview (with discussion). *Test*, **10**: 1–73